Article

new media & society
2019, Vol. 21(3) 562–582

DOI: 10.1177/1461444818801010
journals.sagepub.com/home/nms
SAGE

Equal access to online information? Google's suicide-prevention disparities may amplify a global digital divide

Sebastian Scherr
University of Leuven, Belgium

Mario Haim and Florian Arendt
University of Munich (LMU), Germany

Abstract

Worldwide, people profit from equally accessible online health information via search engines. Therefore, equal access to health information is a global imperative. We studied one specific scenario, in which Google functions as a gatekeeper when people seek suicide-related information using both helpful and harmful suicide-related search terms. To help prevent suicides, Google implemented a "suicide-prevention result" (SPR) at the very top of such search results. While this effort deserves credit, the present investigation compiled evidence that the SPR is not equally displayed to all users. Using a virtual agent-based testing methodology, a set of 3 studies in 11 countries found that the presentation of the SPR varies depending on where people search for suicide-related information. Language is a key factor explaining these differences. Google's algorithms thereby contribute to a global digital divide in online health-information access with possibly lethal consequences. Higher and globally balanced display frequencies are desirable.

Keywords

Agent-based testing, big data, computational methods, digital divide, digital inequality, Google algorithm, Internet search engines, online information seeking, online search behavior, suicide prevention

Corresponding author:
Sebastian Scherr, School for Mass Communication Research, University of Leuven, Parkstraat 45, 3000 Leuven, Belgium.
Email: sebastian.scherr@kuleuven.be

Suicide is a major health problem causing approximately 800,000 annual deaths worldwide—that is one suicide every 40 seconds—and is the second leading cause of death among 15- to 29-year-olds (WHO, 2014, 2017). The social, emotional and economic consequences of suicides are dreadful. Importantly, the pivotal role of the media for suicides *and* suicide prevention is not questioned anymore by suicide prevention experts (see WHO, 2017). In fact, nonfictional and fictional media depictions of suicides grant access to both harmful and protective information (e.g. Till et al., 2017).

Google optimized its search-results page to provide better access to suicide prevention information (Google Official Blog, 2010, November). Consequently, Google implemented a so-called suicide-prevention result (SPR; see Figure 4 in the Appendix 1 for a screenshot), which is prominently displayed at the top of its search results and depicts important online and offline help resources such as country-specific helpline telephone numbers, crisis chat rooms, and websites. For Google, establishing the SPR reflects their self-imposed corporate social responsibility to "do the right thing" (Alphabet Inc., 2017), including its web search service, which in turn potentially grants access to both helpful and harmful information about suicide (e.g. Gunn and Lester, 2013; Thornton et al., 2017). For example, "doing the right thing" can mean to increase the display rates of the SPR by integrating social category information (and related trace evidence) into the algorithms, which can, however, cause outcome bias, or statistical discrimination (see Williams et al., 2018). Relatedly, local data-protection laws and Google's corporate responsibility for data privacy can be in tension with more widely offering the SPR, further optimizing underlying algorithms by analyzing user data, and not jeopardizing the users' trust in the platform (e.g. Dinev and Hart, 2005; Sundar and Marathe, 2010). Google's SPR can be understood as a tailored awareness message that appears exactly when someone googles for a potentially harmful, suicide-related search term—similar to an intervention, that is presented to users who are deemed to be in an acute suicidal state based on their search terms. The SPR presents helpful, potentially protective information *right in the moment* when apparently needed (i.e. when users are looking for suicide-related terms), a beneficial temporal contiguity. Given the high number of Google users worldwide (Internet Live Stats, 2017), this has enormous potential for suicide prevention.

Shedding light on algorithmic decision-making, especially in socially relevant fields such as public health, is crucial in order to understand inequalities, biases, discrimination, and other undesired consequences of technology (Beer, 2017). Importantly, algorithms invisible to the public eye make it difficult to hold companies accountable for their algorithmic decisions (Brake, 2017; Janssen and Kuk, 2016). Hence, the present study contributes to the discussion on algorithm transparency and accountability as well as to technocratic or algorithmic governance in a broader sense (Janssen and Kuk, 2016; Just and Latzer, 2016).

Google does not disclose its algorithmic detection of acute suicidality. Only one previous study from Germany showed that Google's SPR is displayed to a limited number of potentially vulnerable individuals: One-in-ten Google searches for helpful search terms (e.g. "overcoming suicidal thoughts") and only one-in-four searches for harmful search terms (e.g. "best method for suicide") triggered the display of the SPR (Haim et al., 2017). There is hardly any knowledge about the presentation rate of the SPR

worldwide. The present study investigated SPR's display frequency globally. In fact, we focus on the question whether Google contributes to disparities in access to suicide prevention information globally (i.e. unequal presentation rates in different countries and for different languages). This may contribute to a global digital divide, possibly amplifying global gaps in access to health information over time (Norris, 2001).

In the present paper, we first provide a review of research on the media's role in the suicide domain. Second, we theorize on digital divides and relate this line of research with suicide prevention. Third, we present a set of three studies, in which we used agent-based testing methodology as a novel computational research approach to empirically test global health-information disparities linked with the SPR. Finally, we discuss the implications of a suicide prevention-related global digital divide, emphasizing how low SPR display frequencies relate to omitted opportunities to save lives.

Suicide and (new) media

Suicides cluster temporarily and locally, and therefore, media depictions of suicides generally qualify as a resource for suicide role models (Boyce, 2011). Particularly for celebrities and if the news coverage was prominently placed in the media, previous research provides empirical evidence for a subsequent increase in suicide rates—a phenomenon known as "Werther effect" (Niederkrotenthaler et al., 2012; Phillips, 1974; Romer et al., 2006; Stack, 2005). Based on such findings, media outlets are encouraged to avoid detailed sensationalizing suicide coverage, and to present non-sensationalizing suicide reports together with help information, such as helpline telephone numbers (Sudak and Sudak, 2005).

Researchers also accumulated evidence for a reversed, beneficial media effect, labeled as "Papageno effect" (Niederkrotenthaler et al., 2010), that occurs after the media depict people overcoming suicidal crises, and thereby help audiences to overcome suicidal crises themselves (see also Sisask and Värnik, 2012). Both effects were more recently conceptualized as the endpoints of a continuum from harmful (Werther) to helpful (Papageno) effects depending on the mediated suicide depiction, individual characteristics of media users, as well as biopsychosocial and environmental factors (Scherr, 2016; Scherr and Steinleitner, 2015).

Over the past decade, the Internet has been increasingly acknowledged as a tool for suicide prevention (Biddle et al., 2008; Gunn and Lester, 2013; Mehlum, 2000). Particularly, the importance of an easy access to health information was highlighted frequently, since many suicides are preventable (Wasserman, 2016b) if vulnerable people in a suicidal crisis are granted access to adequate help information (Lester and Rogers, 2012). Search engines were recently conceptualized as a gate to access tailor-made, suicide-related, health information (Arendt and Scherr, 2016; Shoemaker and Vos, 2009). As such, different online platforms and environments received much attention within research on Werther and Papageno effects, since they can prevent users from dying by suicide, but have also been reported to make suicide-related information easily accessible with potentially detrimental effects (see Collings and Niederkrotenthaler, 2012; Daine et al., 2013). Notably, at a societal level, suicides can evoke specific online search

behavior (Arendt and Scherr, 2017) that correlates with actual suicide rates (Arendt and Scherr, 2016).

Given this importance of online environments to health information, the quality of such information that suicidal individuals potentially have access to when searching for information about suicide methods remains crucial (Biddle et al., 2008, 2016; Kemp and Collings, 2011; Recupero et al., 2008; Thornton et al., 2017; Till and Niederkrotenthaler, 2014). For instance, Till and Niederkrotenthaler (2014) analyzed websites that were retrieved from searches for the term suicide, method-related search terms, and help-related search terms using search engines in the United States and Austria. While two thirds of all retrieved results were protective websites, their quality depended substantially on the search terms: Especially method-related searches contained fewer protective and more harmful characteristics. Similarly, Thornton et al. (2017) compiled a list of suicide-related search terms which they typed into the Google search in Australia, and found that 69% of the relevant websites as accessed via Google were deemed to be suicide-preventive and only 2% were explicitly pro-suicide. The authors also noted that Google displayed its SPR only after 8 of the 26 search terms.

Google's SPR as a nudge—but for everyone?

Only recently, Haim et al. (2017) used virtual agents and focused on this latter finding more systematically in the German context. The study modeled online search behavior of 1200 virtual agents using an agent-based testing methodology and found that Google's SPR showed up in 11% of the cases when helpful search terms were used, and after 25% of harmful searches. Hence, based on algorithmic decisions, Google apparently shows the SPR only rarely to its German users.

According to a *New York Times* article (Cohen, 2010), a Google employee first raised the company's awareness of the issue that potentially harmful information related to suicide (e.g. method-related information on how to kill oneself) are easily accessible via search requests that might also be an indicator that a user is currently facing a substantial suicidal crisis. Consequently, Google implemented a box on top of the search results when potentially harmful, suicide-related search terms are used. The box originally containing the National Suicide Prevention Lifeline and was available in 14 countries (Australia, Belgium, France, Germany, Hungary, Italy, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the United Kingdom, and the United States). In November 2010, Google estimated the increase in hotline calls in the United States to be 9% in the months after having implemented the SPR (Google Official Blog, 2010, November), and planned to extend it to more countries in the future.

The relevance of a SPR including a reference to counseling services cannot be underestimated. While early studies on beneficial effects of telephone hotlines in overcoming a suicidal crisis showed only mixed support (see Lester, 1997), more stringent field studies later yielded more promising results. These results show that callers' suicidal ideation, suicidal urgency, intent to die, hopelessness, and psychological pain could be reduced during telephone counseling, and even continued to decrease in the weeks after the call (Gould et al., 2017; Knox and Bossarte, 2009). Likewise, the Department of Veterans Affairs' suicide prevention hotline and chat services proved effective when raising public awareness about these services significantly increased their usage (see Knox

and Bossarte, 2009). More specifically, they stopped 79.6% of the clients from killing themselves and kept 9-in-10 (90.6%) people safe (Gould et al., 2017). An explanation for these findings is arguably that the SPR not only provides knowledge about specific suicide help resources such as the National Suicide Prevention Lifeline and other services (e.g. crisis text line, Trans Lifeline), but also functions as a prompt for their existence during suicide crises.

Moreover, the display frequency of the SPR is automatized and most likely responds to individual search queries, this is a novel algorithmically curated and data-driven perspective on suicide prevention. In contrast to reports on suicides in traditional mass media, which have been conceptualized to potentially help a wide array of suicidal people by pointing out means to overcome a suicidal crisis (Niederkrotenthaler et al., 2010), the SPR directly reaches out to suicidal people right during purposeful information-seeking. Given a restrictive suicidal tunnel vision with suicidal plans on top of the minds of suicidal people (Wasserman, 2016a, 2016b), it is important to note that this moment of being exposed to preventive information is not only characterized by cognitive restriction but also by high ambivalence about life and death (Wasserman, 2016b). As such, even minor supportive instances can prevent people from death in this phase. Hence, the SPR theoretically reaches ambivalent suicidal people in the right moment.

In health contexts, the effectiveness of prompts (or primes, or nudges) has been shown (e.g. Kennedy et al., 2004; Thaler and Sunstein, 2008). For instance, calls to the national US STD and AIDS hotline services increased heavily shortly after their telephone number was displayed after two episodes of a soap opera (Kennedy et al., 2004). Beyond fictional media, hotline calls were shown to increase after campaign messages (Knox and Bossarte, 2009). Hence, Google's own data and previous health communication research is consistent with the assumption of SPR eliciting a beneficial effect.

From what is known about algorithmic biases and its relation to digital inequalities, companies do not develop algorithms to be imbalanced, but often fail to validate and improve them after their implementation for three reasons (Williams et al., 2018): First, given that an initial implementation already yields improvements as compared to a situation where no algorithms were in place, validation and further improvements require resources without compelling cause. For example, ever since the SPR has been implemented, an increase in phone calls to the national helpline has been reported (Google Official Blog, 2010, November). For Google, validating and improving the display rate of the SPR from this point onward is time-consuming without further reward of public appreciation. Second, validations and improvements require sensitive individual data, thus touching a company's responsibilities for data protection and privacy. Therefore, implementing "good" algorithms, such as the SPR, may express corporate social responsibility, whereas maintenance and optimization require the company to be responsible regarding data protection—a corporate dilemma. In the case of Google, the implementation of potentially suicide-preventive algorithms reflects the company's corporate social responsibility (Gunn and Lester, 2013). Yet, optimizing the algorithmic SPR decision-making might be restricted by Google's data protection policy. Third, modern self-learning algorithms could in fact produce country-specific differences, if they are not explicitly designed to avoid disparities (Williams et al., 2018). Of course, only Google itself can definitely explain how the decision-making behind the SPR works, but remained irresponsive to our requests to disclose.

Showing up some health information only to some users describes the very problem of a "digital divide" (Norris, 2001) or digital inequalities (DiMaggio and Hargittai, 2001). The two terms originally described socioeconomic disparities regarding the physical access to computers and the Internet. However, since Internet-access gaps narrowed substantially over the past decade (Pew Research Center, 2009), the focus shifted toward second-level digital divides (i.e. Internet usage gaps) between social or racial/ethnic groups (Hargittai and Hinnant, 2008; Zillien and Hargittai, 2009), which have been conceptualized as complex and non-static processes with the tendency to deepen further, independent of physical access (Van Dijk, 2005; Van Dijk and Hacker, 2003). The more recent focus on Internet-use inequalities also includes specific environments, such as search engines and social media, and their health effects (see, for example, Kontos et al., 2010), and relates digital divides to content production and geography (Robinson, 2009; Robinson et al., 2015; Schradie, 2011), and go beyond the familiar idea of the "digital divide" (Halford and Savage, 2010).

Van Dijk (2017) differentiates digital inequalities as either (1) theorizing about what brings people to accept, access, and use a new technology, (2) explanations for differences between groups based on their social network position and the accumulation of capital, or (3) an effect chain from personal and positional inequalities, over the distribution of resources, to the characteristics, and access to a technology, and its use. Relatedly, based on anecdotal evidence from different countries, Warschauer (2002) argued that even if equal technological access was granted, multiple restrictive layers remain, and, for example, language, media literacy or other structural influences maintain digital divides (Graham, 2011; Warschauer, 2002). Consequently, digital inequalities may also arise from unequal access to a technology depending on geolocation—an assumption that has yet been barely explored (see Van Dijk, 2017).

Given that Google is a US company and has implemented the SPR at first in the United States almost 10 years ago, we assumed that the implementation process in different countries around the globe would take place at different speeds and would depend on legal restrictions in the countries, thereby inducing relevant health-information disparities related to the display frequency of the SPR. We conducted a series of three studies in which we investigated the display frequency of the SPR at the global scale: Study 1 replicates the pioneer work in this area (Haim et al., 2017) and extends it onto the US context. We describe the methodological approach in detail here, since Study 2 (extension to global scale) and Study 3 (within-country differences based on language) use the same methodology. Each study therefore makes a unique contribution.

Study 1: replication and extension of previous work

As previously mentioned, prior research on the implementation of Google's SPR in Germany showed that display frequencies reach 25% of the cases when searching for harmful terms (Haim et al., 2017). Due to Google's US origin, we hypothesized substantial differences in the display frequency between the United States and Germany. In fact, we assumed that the display frequency would be substantially higher in the United States:

H1. The display rates of Google's SPR in the United States are higher than in Germany.

Method

For methodological comparability, we chose to use Amazon Web Services® (AWS). Amazon allows to specifically choose in which location a server should be set up, and thus offers country-specific IP addresses that are neither associated with private households nor universities. It therefore allows us to investigate search behavior as if ordinary citizens were performing the searches in the respective country. To control for any individual server influences, we set up two servers for each setting (i.e. for every language in every country).

Search terms. All agents were programmed to repeatedly search for terms from a list of either suicide-related helpful, suicide-related harmful, or suicide-unrelated search terms. The original list of suicide-related terms was suggested elsewhere and includes between 10 and 20 terms per group (Biddle et al., 2008). The terms unrelated to suicide (control group) were 28 nouns, verbs, or adjectives with a mean arousal value of between 2.7 and 2.8 taken from the well-established Berlin Affective Word List (BAWL; see Võ et al., 2009). All search terms can be found in the open access online supplement to this article (see Table S1).

Replication software script. The software was based on CasperJS, a NodeJS-based server-sided framework for scripted web-surfing activities. All scripts are fully available online (https://github.com/MarHai/ScrapeBot/releases/tag/v1.1). In order to increase the study's external validity, we used several virtual agents within each setting (i.e. for every language in every country). Hence, for each setting, the software repeatedly chose one random virtual agent out of the three agent groups (i.e. helpful, harmful, control group), which then randomly selected one search query from the agent group's word list in the language spoken in the respective country. It then set up a new browser environment without any prior cookies or browsing history, navigated to https://www.google.[ccTLD], performed one search, checked whether it had been redirected to Google's result page, and recorded whether the SPR was displayed or not. An inspection of randomly taken screenshots did not indicate any malfunction.

All data have been made publicly available at https://doi.org/10.7910/DVN/PCRG1D.

Measures

It was tested whether and how often the SPR shows up in different countries and for different languages. The main outcome was the display frequency of the Google SPR to users when (a) suicide-related helpful, (b) suicide-related harmful, or (c) suicide-unrelated search terms (the control condition) were queried in a Google search.

Statistical analysis

We used the absolute number of displays of the SPR and present 95% confidence intervals based on 1000 percentile-bootstrapped replication samples to provide nonparametric

confidence intervals for our absolute number of displays. These allow us to visually inspect and test for significant differences regarding how often the SPR was displayed and can be found in the figures.

Procedure

Study 1 was conducted in January 2017. We used four servers, two in Germany (Frankfurt), and two in the United States (East and West coast), as provided by AWS. On each server, we installed three groups of $n = 20$ virtual agents per group. All agents simulated regular Google search behavior. In total, we modeled the online search behaviors of $N = 4$ servers $\times$ 3 agent groups $\times$ 20 virtual agents $= 240$ virtual agents using agent-based testing over the course of 10 days.

In both countries, our software repeatedly chose one random virtual agent that then randomly selected one search query from the agent's word list, navigated to https://www.google.[ccTLD], performed the search, and recorded whether the SPR was displayed or not. To note, virtual agents imitate real human searches. All virtual agents repeatedly navigated to their respective local Google website, searched for a predefined list of suicide-related helpful, suicide-related harmful, or suicide-unrelated search terms, and then recorded for 10 consecutive days whether the SPR was displayed or not. Over the course of 10 days, this setup yielded 120,910 search queries in Germany and 110,417 queries in the United States.

Results

We found that in contrast to the United States, where the SPR was displayed in the aftermath of approximately 94% of queries for harmful and approximately 34% of the helpful suicide-related search terms, in Germany, the SPR was only displayed in about 23% of the searches for harmful and in around 8% of the helpful search terms (see Figure 1 including 95% confidence bands of point estimates). Importantly, as can be seen in Figure 1, these display rates stayed approximately at the same level over time.

Discussion

We found a large and mainly stable disparity between the United States and Germany, consistent with a digital divide in Study 1. From the perspective of suicide prevention, this finding is highly disturbing given that the geographic location of the person who searches for suicide-related information determines an individual's odds of coming across information about where to find help during a suicidal crisis.

Study 2: global extension of study 1

Given that the search engine Google operates globally, digital divides in access to suicide prevention information as found in Study 1 may be present on a global scale. Therefore,

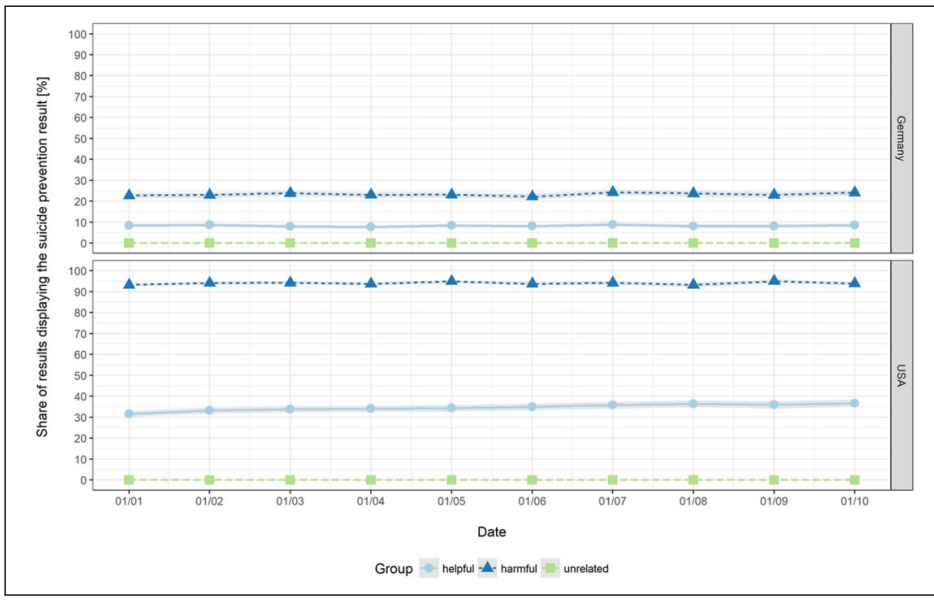

Figure 1. Relative display frequencies for the Google suicide-prevention result (SPR) after inputting suicide-related helpful, suicide-related harmful, or suicide-unrelated (i.e. control) search terms in Germany and in the United States in early January 2017. The figure displays the percentages including the 95% confidence bands based on percentile bootstrapping with 1000 replications.

Study 2 explored the display rates of Google's SPR globally. We formulated the following research question:

RQ1. In how far do the display rates of Google's SPR differ across the globe?

Method

We extended the methodological approach as described for Study 1 to a global scale by using two or more servers in all countries where AWS was available during early July 2017. We used the same search terms as in Study 1, included further suggestions for helpful or harmful search queries from local suicide experts, and had the search terms translated by these experts and native speakers into each country's main spoken language.[1] The search queries, again, were repeatedly performed by virtual agents, this time over the course of 7 days, resulting in a total of 1,169,682 search queries (see also Table *S2*). Our sample included 20 servers in Australia, Brazil, Canada, Germany, India, Ireland, Japan, Singapore, South Korea, and the United Kingdom (two server each) plus four servers in the United States. Overall, and in accordance with Study 1, we employed $N = 24$ servers $\times$ 3 agent groups $\times$ 20 virtual agents $= 1440$ virtual agents using the same agent-based testing methodology.

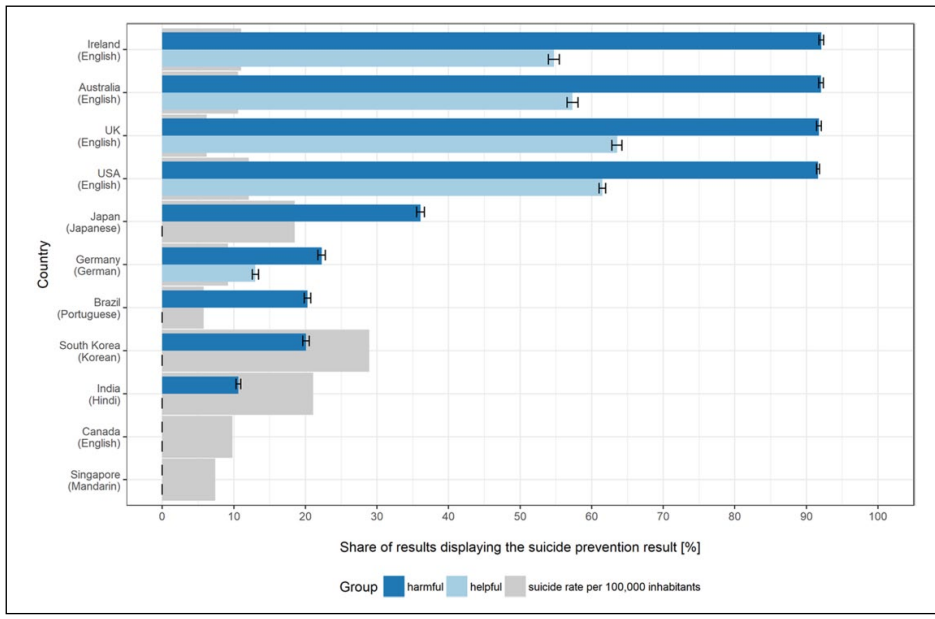

Figure 2. The display percentages for the Google suicide-prevention result (SPR) after inputting suicide-related helpful or harmful search terms, or search terms unrelated to suicide across the globe as of early July 2017. The figure displays the percentages including the 95% confidence intervals based on percentile bootstrapping with 1000 replications. The absolute number of performed searches and the relative display rates of the SPR can be obtained from the Online Supplement from Table S2. The gray bars depict the suicide rates per 100,000 in the countries from 2012 (WHO, 2014).

Results

Study 2 shows striking differences between countries with regard to how often the SPR was displayed (see Figure 2 including the 95% confidence intervals of point estimates and Table S2 in the Online Supplement for the exact figures). The countries in which the SPR was presented most-often after suicide-related harmful search terms were used include Australia (92%), Ireland (92%), the United Kingdom (92%), and the United States (92%). However, in Japan (36%), Germany (22%), Brazil (20%), South Korea (20%), and India (11%), the SPR was shown at substantially lower rates. Strikingly, in Canada and Singapore, the SPR was never displayed, even though users requested information that was clearly linked to potential self-harm. When users searched for information about how to overcome a suicidal crisis (i.e. helpful search terms), the help-directing SPR only showed up in the United Kingdom (64%), the United States (61%), Australia (57%), Ireland (55%), and Germany (13%).

However, we also found no significant differences, indicated by overlapping 95% confidence intervals, for harmful search terms in Ireland, Australia, the United Kingdom, and the United States, with a display rate of 92%, or for Brazil and South Korea, where

the SPR was displayed after 20% of potentially harmful searches. In contrast to our sample of two servers (one on each US coast) in Study 1, we used all four US server locations available through Amazon in Study 2. We therefore believe that the slight variance regarding the frequency with which the SPR was displayed (Study 2: 92% vs Study 1: 94%) after suicide-related harmful search terms were inputted is a result of the methodological extension performed in Study 2. In line with Study 1, variation over time in each country was not substantial.

Discussion

Study 2 provides additional evidence for health information disparities. Supplementing the findings about a digital divide between the United States and Germany found in Study 1, Study 2 offers multi-country evidence consistent with a global digital divide. The question emerges, which factors might be able to explain these disparities. Overall, the findings of Study 2 suggest that Google seems to prioritize English searches over other languages. Some other factors that potentially influence the display of the SPR such as country-specific top level domain, public/institutional vs private IP address, browsing history, activities within the Google results, or page loading speed have already been ruled out (Haim et al., 2017). Thus, Study 3 was conducted in order to address particularly the role of language (Warschauer, 2002).

Study 3: within-country differences related to language

Since digital divides can be categorized by attributes related to geography (see Study 1 and Study 2) and language (see Hilbert, 2011; Warschauer, 2002), we additionally investigated display frequencies in countries, in which more than one official language is actively spoken by the population. In such countries, the assumed digital divide is of exceptional importance as it would classify access to health information not only based on national but rather on individual predispositions related to language use.

> *RQ2*. In how far do the display rates of Google's SPR differ within countries where more than one official language is spoken?

Method

Given that some countries have multiple official languages, we conducted Study 3 in Singapore focusing on two languages, and in India focusing on three languages. In Singapore, English is the official main language in schools, for administration, and for international business, but Mandarin is the most common language spoken at home. In India, Hindi is the most widely spoken language, but English is important within the educational and government sectors. Relatedly, Telugu is the third most widely spoken native language and one of four languages in India that have an official status in more than one state.

Four servers were located in Singapore, two of which employed programmed virtual agents using Mandarin search queries, whereas the other two used search queries in

English. In India, six servers were based in Mumbai, with virtual agents using the same search terms in either English, Hindi, or Telugu (two servers each). In total, we used $N = 10$ servers $\times$ 3 agent groups $\times$ 20 virtual agents $= 600$ virtual agents using agent-based testing.

We conducted this third study over the course of 4 days in mid-July 2017. Using the same search terms as in Study 1 and Study 2, virtual agents repeatedly performed random search queries, resulting in a total of 202,309 search queries (see Table S3 in the Online Supplement for the number of performed searches).

Results

The results, depicted in Figure 3, show that when suicide-related harmful search terms were typed in English, users saw Google's SPR in 92% of the searches in Singapore and 91% of the searches in India (95% confidence bands of point estimates depicted in Figure 3). However, when users in India went online to look for potentially harmful search terms in Hindi, they experienced massively lower odds of being exposed to the SPR, as it was displayed in only 10% of the searches. Strikingly, potentially harmful searches performed in Telugu (India) or Mandarin (Singapore) did not yield even one display of the SPR.

Furthermore, when users searched for helpful terms in English, the SPR showed up in 57% of the cases in India and in 47% of the cases in Singapore, but it was never shown in Hindi, Telugu, or Mandarin. When search terms unrelated to suicide were used, the SPR was never displayed.

Discussion

Analyses indicate that language is a factor that explained SPR's display frequencies within a given country as it had been theorized earlier (e.g. Warschauer, 2002). The differences in display frequency were substantial. Apparently, it hugely mattered in which language a given individual searches for suicide-related information. English searches consistently yielded more SPR displays than searches in other official languages. Hence, some users received more suicide prevention information than others despite the fact that everybody used the same search terms.

General discussion

Suicide is a tremendous but preventable global health problem, for which the media are considered a pivotal risk and prevention factor (WHO, 2014, 2017). There is paramount evidence that both fictional and nonfictional media contribute to imitative (copycat) suicides (Gould et al., 2014; Niederkrotenthaler et al., 2012; Stack, 2005, 2009) and solid yet still growing ground support for the media's preventive role during suicidal crises (Niederkrotenthaler et al., 2010; Sisask and Värnik, 2012), especially over the Internet (Biddle et al., 2016; Scherr and Reinemann, 2016; Thornton et al., 2017; Till et al., 2017; Till and Niederkrotenthaler, 2014). Acknowledging the Internet's role in suicide prevention, Google implemented the SPR—an info box with references to

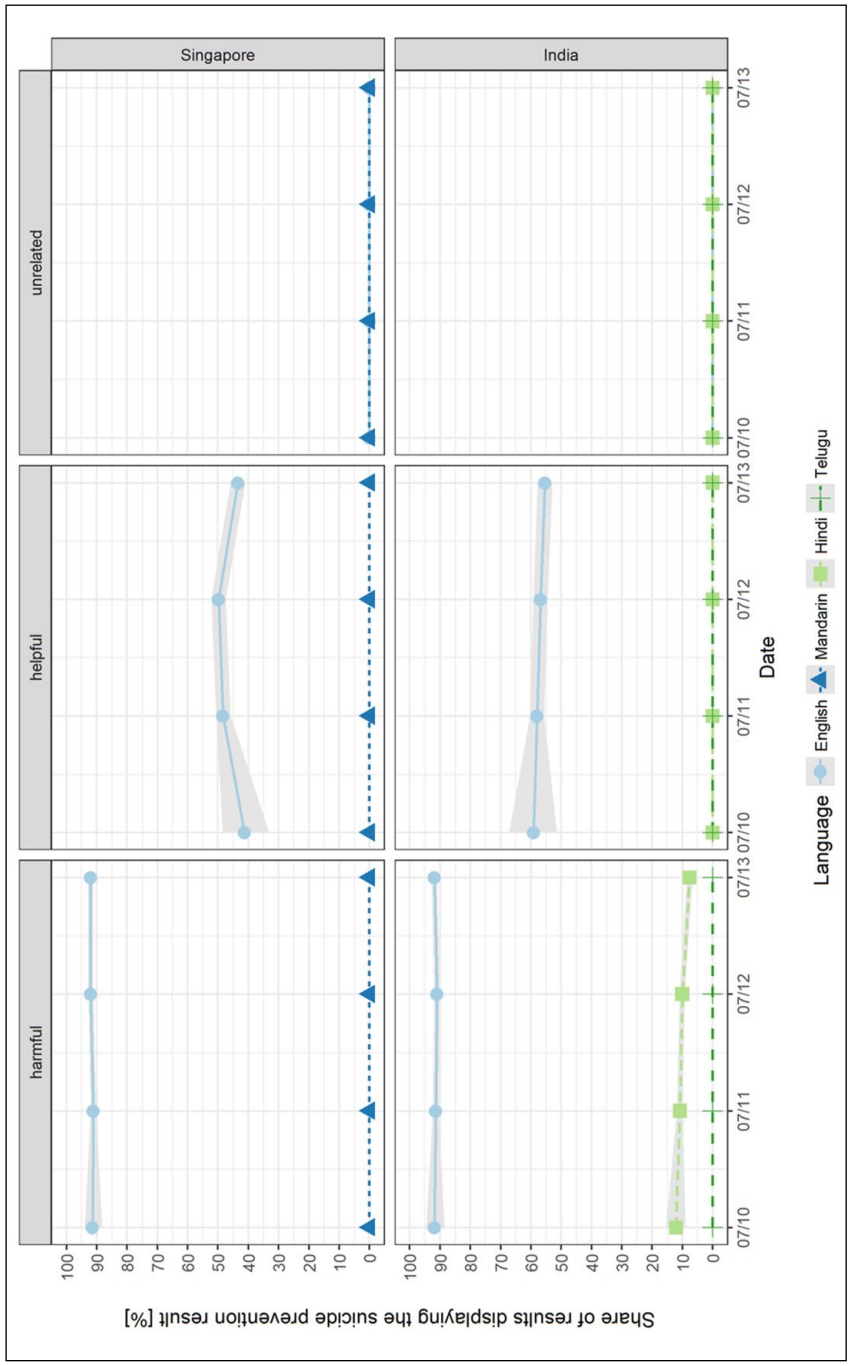

Figure 3. The relative display frequencies for the Google suicide-prevention result (SPR) after inputting suicide-related helpful or harmful search terms, or search terms unrelated to suicide in the different key spoken languages in Singapore (above) and India (below) in mid-July 2017. The figure displays the percentages including the 95% confidence bands based on percentile bootstrapping with 1000 replications. The absolute number of performed searches and the relative display rates of the SPR can be obtained in the Online Supplement from Table S3.

counseling services and other web resources that help people overcome acute suicidal crises.

In order to explore suicide-related health information disparities, we conducted a set of three consecutive studies. Using agent-based testing, we automatically searched for three bags of search terms with suicide-related and potentially harmful (e.g. "how to commit suicide"), potentially helpful (e.g. "how to overcome a suicidal crisis"), or unrelated words that were suggested by previous research. In total, our three studies produced almost 1.6 million Google search queries where, for each search, we stored whether Google's SPR was displayed or not.

In the first two studies, we found considerable disparities in the display rates of Google's SPR between the United States and Germany (Study 1). Extending this finding, Study 2 offers evidence consistent with the notion of a global digital divide. Study 3 focused on language, a possible underlying mechanism, by looking at countries where English and other official languages are spoken (India, Singapore). Again, we found large inequalities insofar as the parts of these populations that search in languages other than English face dramatically lower odds of being exposed to suicide-preventive help resources. Apparently, language is an important factor that explains information disparities produced by search engines. Thus, differences in the display frequency of the SPR may lead to substantial health disparities.

Policy implications for suicide prevention in times of the Internet

Ultimately, suicide prevention should aim for maximizing the frequency of the SPR being displayed to vulnerable users, especially in presumably high-risk situations (i.e. for harmful search terms), but also when people actively search for help (i.e. for helpful search terms). Withholding such information can miss the opportunity to save lives, while false positives (i.e. showing the SPR to users who do not necessarily need it) do not cause any health-related harm.

We acknowledge that Internet companies fear that a "too-well-adjusted" algorithm might produce many false-positives, evoking emotional discomfort or privacy concerns, possibly leading to more negative attitudes toward the search engine or its advertised content. Ultimately, a displayed SPR reduces available advertising space. Providers of search engines have a social responsibility. Many false-positives count less than one missed opportunity to save the life of a vulnerable individual. While we acknowledge the complexity of algorithms (Lazer, 2015), they are adjustable and should be constantly tested and improved (Lazer et al., 2014). The lists of search queries used in the present studies are provided as an Open Access Online Supplement and are meant to help in this optimization process.

Moreover, our studies have implications regarding the external validity of findings. For instance, a recent experimental study randomly assigned participants to read a news article that contained preventive resources and psychoeducational information or not (see Williams and Witte, 2017). Knowledge about suicide, suicide warning signs, and attitudes toward suicide and help-seeking were largely unaffected by the manipulation. Apparently, the prevention information was not enough systematically elaborated to produce effects on participants. Tailored prevention information as provided by the SPR

should be elaborated more systematically in the moment when people are undergoing a suicidal crisis. What kind of messages influence attitudes toward suicide, suicide knowledge, and help-seeking in the most beneficial way? Apparently, this research question needs to be tested in the future as well.

Implications for research on digital divides

Discussions about digital divides should start to question their premises that all users have the same chance to see the same content on the Internet and that the term "algorithm" necessarily involves highly complex automated "reasoning." We found systematic, algorithm-based health-information disparities across the globe. However, the majority of studies in this area follows a methodologically individualistic perspective and discussions include disparities in access (e.g. net neutrality), disparities in information quality and use patterns through personal or (network) positional inequalities (e.g. filter bubbles), resource distributions, or technological characteristics (see UNICEF, 2017; Van Dijk, 2017). Importantly, discussions about these latter aspects usually assume that all users have access to the same content. However, the phenomenon is different, if access to the same content is not granted to everybody as indicated in the present set of studies. This has been implicitly addressed by Hilbert (2011) who characterized digital divides by the technology, the subjects affected by the divide, and their attributes: apparently, also combinations of these characteristics, that is, algorithm-based disparities in web content across countries that can be explained by language, contribute to digital divides. This notion also touches the discussion about "language-as-data" and the competing motives of Internet companies such as Google or Apple and Internet users as put forward by Thornton (2017). Regarding the SPR, Google not only provides the technology, but also limits its access to users based on individual attributes such as language. Hence, we found evidence for *differential access to information* (see also Duff, 2012). Research has not yet addressed this substantially; however, differential access to information raises important societal questions (e.g. on net neutrality).

Relatedly, the effects thereof have been barely addressed. To our knowledge, the most comprehensive evidence for the effects of unequal access to Internet information stems from a survey that tackled the (dis)advantages of having (v. not having) access to different kinds of information online (Van Deursen et al., 2014). However, information in this study was generally available. Our findings suggest that in the specific area of suicide prevention, Google's algorithms—and Google itself, respectively—decide on who gets to see information about help resources in a suicidal crisis. This decision is algorithm-based and biased toward English-speaking populations.

Limitations

Our studies are subject to several important limitations. First, given that we relied on Amazon's server infrastructure, specific idiosyncrasies might relate to our findings (e.g. How were the IP addresses resolved? Was there any forwarding?). Alternatives would be to run identical servers on all continents, for which we did not have the means, or to have Google helping to explain how the SPR is programmed. However, all of our repeated requests to obtain information from Google remained unanswered. Future studies, ideally

with access to a wider range of countries *and* a comparable server network, should extend the approach presented here, also with regard to the used search terms. Second, in order to have a mean for comparison, we simply used the same list of search terms as suggested by prior studies (Biddle et al., 2008; Haim et al., 2017). This list may not be complete. Third, the present study involved several tradeoff decisions, which again resulted in our reliance on the Amazon infrastructure, and using it at least on all available continents (as of mid-2017). Thereby, an important omission is for instance the African continent, where no such services were available. Finally, even though accurate for the present study, agent-based testing has important methodological limitations (Bruch and Atwell, 2015). There are no norms about what defines a virtual agent, about different types of agent architecture and agent organization, and about in how far the agents are allowed to interact with each other. Therefore, it is difficult to compare our methodology to those of other studies with different premises. As with actual human search behavior, it remains widely unclear, if and how Google's algorithms respond to the premises set for our virtual agents. Conceptually, we can only speak to a lack of access to suicide prevention information as a contributing factor to digital divides resulting from the users' language and location.

Conclusion

For some of its scholarly services, Google uses a quote by Isaac Newton as its motto, stating that it "stands on the shoulders of giants," which refers to the cumulative power of knowledge gain. However, our set of three coherent studies shows that not everybody is standing on the shoulders of the same giant regarding the display frequency of the SPR. Google deserves credit for having integrated this feature into its search results. Unfortunately, helpful information about where to find help in suicidal crises is only presented to a limited number of potentially vulnerable individuals across the globe with disturbing differences between—and even within—countries. Importantly, during suicidal crises, the cost of a "false alarm" is substantially lower than the cost of a "miss," which may ultimately result in the death of a vulnerable member of our global society. Access to health information, especially during suicidal crises, should not depend on the country where the online search takes place, nor on a user's language. By adjusting Google's algorithmic decision-making process, the existing global digital health divide can be narrowed and lives can be saved at virtually low cost.

Authors' Note

Raw data are available from https://doi.org/10.7910/DVN/PCRG1D.

Acknowledgements

Sebastian Scherr, School for Mass Communication Research, University of Leuven; Mario Haim, Department of Communication Science and Media Research, University of Munich (LMU); Florian Arendt, Department of Communication Science and Media Research, University of Munich (LMU).

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. We are sincerely thankful to Sateesh Babu, Rui C. Campos, Rui Qi Choo, Gillian Green, Eunkyung Ko, Ownkyeong Lee, Michael Prieler, Dan Reidenberg, and Hajime Sueki for helping us with the translations of the search terms into their language. The languages used in each country were (in alphabetical order) English (Australia, Canada, Ireland, United Kingdom, United States), German (Germany), Hindi (India), Japanese (Japan), Korean (South Korea), Mandarin (Singapore), and Portuguese (Brazil).

Supplemental Material

Supplemental material for this article is available online.

ORCID iD

Sebastian Scherr https://orcid.org/0000-0003-4730-1575

References

Alphabet Inc (2017) Alphabet investor relations: code of conduct. Available at: https://abc.xyz/investor/other/code-of-conduct.html

Arendt F and Scherr S (2016) Optimizing online suicide prevention: a search engine-based tailored approach. *Health Communication 32*(11): 1403–1408.

Arendt F and Scherr S (2017) The impact of a highly publicized celebrity suicide on suicide-related online information seeking. *Crisis: The Journal of Crisis Intervention and Suicide Prevention 38*(3): 207–209.

Beer D (2017) The social power of algorithms. *Information, Communication & Society* 20: 1–13.

Biddle L, Derges J, Mars B, et al. (2016) Suicide and the Internet: changes in the accessibility of suicide-related information between 2007 and 2014. *Journal of Affective Disorders* 190: 370–375.

Biddle L, Donovan J, Hawton K, et al. (2008) Suicide and the internet. *BMJ* 336: 800–802.

Boyce N (2011) Suicide clusters: the undiscovered country. *The Lancet* 378: 1452.

Brake DR (2017) The invisible hand of the unaccountable algorithm: how Google, Facebook and other tech companies are changing journalism. In: Tong J, Lo S-H (eds) *Digital Technology and Journalism: An International Comparative Perspective*. Cham: Palgrave Macmillan, pp. 25–46.

Bruch E and Atwell J (2015) Agent-based models in empirical social research. *Sociological Methods & Research* 44: 186–221.

Cohen N (2010) "Suicide" query prompts Google to offer hotline. Available at: http://www.nytimes.com/2010/04/05/technology/05google.html

Collings S and Niederkrotenthaler T (2012) Suicide prevention and emergent media: surfing the opportunity. *Crisis: The Journal of Crisis Intervention and Suicide Prevention* 33: 1–4.

Daine K, Hawton K, Singaravelu V, et al. (2013) The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. *PLoS ONE* 8: e77555.

DiMaggio P and Hargittai E (2001) *From the "Digital Divide" to 'Digital Inequality': Studying Internet Use as Penetration Increases*. Princeton, NJ: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University.

Dinev T and Hart P (2005) Internet privacy concerns and social awareness as determinants of intention to transact. *International Journal of Electronic Commerce* 10: 7–29.

Duff AS (2012) *A normative theory of the information society*. New York: Routledge.

Google Official Blog (2010) Helping you find emergency information when you need it. November. Available at: https://googleblog.blogspot.de/2010/11/helping-you-find-emergency-information.html

Gould MS, Kleinman M, Lake AM, et al. (2014) Newspaper coverage of suicide and initiation of suicide clusters in teenagers in the USA, 1988–96: a retrospective, population-based, case-control study. *The Lancet Psychiatry* 1: 34–43.

Gould MS, Lake AM, Galfalvy H, et al. (2017) Follow-up with callers to the national suicide prevention lifeline: evaluation of callers' perceptions of care. *Suicide and Life-Threatening Behavior* 48: 75–86.

Graham M (2011) Time machines and virtual portals: the spatialities of the digital divide. *Progress in Development Studies* 11: 211–227.

Gunn JF and Lester D (2013) Using google searches on the internet to monitor suicidal behavior. *Journal of Affective Disorders* 148: 411–412.

Haim M, Arendt F and Scherr S (2017) Abyss or shelter? On the relevance of web search engines' search results when people google for suicide. *Health Communication* 32: 253–258.

Halford S and Savage M (2010) Reconceptualizing digital social inequality. *Information, Communication & Society* 13: 937–955.

Hargittai E and Hinnant A (2008) Digital inequality: differences in young adults' use of the Internet. *Communication Research* 35: 602–621.

Hilbert M (2011) The end justifies the definition: the manifold outlooks on the digital divide and their practical usefulness for policy-making. *Telecommunications Policy* 35: 715–736.

Internet Live Stats (2017) Google search statistics. Available at: http://www.internetlivestats.com/google-search-statistics/

Janssen M and Kuk G (2016) The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly* 33: 371–377.

Just N and Latzer M (2016) Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society* 39: 238–258.

Kemp CG and Collings SC (2011) Hyperlinked suicide: assessing the prominence and accessibility of suicide websites. *Crisis: The Journal of Crisis Intervention and Suicide Prevention* 32: 143–151.

Kennedy MG, O'Leary A, Beck V, et al. (2004) Increases in calls to the CDC national STD and AIDS hotline following AIDS-related episodes in a soap opera. *Journal of Communication* 54: 287–301.

Knox KL and Bossarte RM (2009) Suicide prevention research—enabling activities funded through VA's office of mental health. In: *Proceedings of the 2009 Department of Defense/Veterans Administration suicide prevention conference*, 12–15 January 2009. San Antonio, TX.

Kontos EZ, Emmons KM, Puleo E, et al. (2010) Communication inequalities and public health implications of adult social networking site use in the United States. *Journal of Health Communication* 15: 216–235.

Lazer D (2015) The rise of the social algorithm. *Science* 348: 1090–1091.

Lazer D, Kennedy R, King G, et al. (2014) The parable of Google Flu: traps in big data analysis. *Science* 343: 1203–1205.

Lester D (1997) The effectiveness of suicide prevention centers: a review. *Suicide and Life-Threatening Behavior* 27: 304–310.

Lester D and Rogers JR (2012) *Crisis intervention and counseling by telephone and the internet*. Springfield, IL: Charles C Thomas.

Mehlum L (2000) The internet, suicide, and suicide prevention. *Crisis: The Journal of Crisis Intervention and Suicide Prevention* 21: 186–188.

Niederkrotenthaler T, Fu K-W, Yip PSF, et al. (2012) Changes in suicide rates following media reports on celebrity suicide: a meta-analysis. *Journal of Epidemiology and Community Health* 66: 1037–1042.

Niederkrotenthaler T, Voracek M, Herberth A, et al. (2010) Role of media reports in completed and prevented suicide: Werther v. Papageno effects. *British Journal of Psychiatry* 197: 234–243.

Norris P (2001) *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. New York: Cambridge University Press.

Pew Research Center (2009) Generations online in 2009. Available at: http://www.pewinternet.org/2009/01/28/generations-online-in-2009/

Phillips DP (1974) The influence of suggestion on suicide: substantive and theoretical implications of the Werther effect. *American Sociological Review* 39: 340–354.

Recupero PR, Harms SE and Noble JM (2008) Googling suicide: surfing for suicide information on the Internet. *Journal of Clinical Psychiatry* 69: 878–888.

Robinson L (2009) A taste for the necessary. *Information, Communication & Society* 12: 488–507.

Robinson L, Cotten SR, Ono H, et al. (2015) Digital inequalities and why they matter. *Information, Communication & Society* 18: 569–582.

Romer D, Jamieson PE and Jamieson KH (2006) Are news reports of suicide contagious? A stringent test in six U.S. cities. *Journal of Communication* 56: 253–270.

Scherr S (2016) *Depression – Medien – Suizid: Zur empirischen Relevanz von Depressionen und Medien für die Suizidalität* [Depression – Media – Suicide: On the empirical relevance of depression and media for suicidality]. Wiesbaden: Springer VS.

Scherr S and Steinleitner A (2015) Zwischen dem Werther- und Papageno-Effekt [Between Werther and Papageno effects]. *Nervenarzt* 86(5): 557–565.

Scherr S and Reinemann C (2016) First do no harm: Cross-sectional and longitudinal evidence for the impact of individual suicidality on the use of online health forums and support groups. *Computers in Human Behavior* 61: 80–88.

Schradie J (2011) The digital production gap: the digital divide and Web 2.0 collide. *Poetics* 39: 145–168.

Shoemaker PJ and Vos T (2009) *Gatekeeping theory*. New York: Routledge.

Sisask M and Värnik A (2012) Media roles in suicide prevention: a systematic review. *International Journal of Environmental Research and Public Health* 9: 123–138.

Stack S (2005) Suicide in the media: a quantitative review of studies based on nonfictional stories. *Suicide and Life-Threatening Behavior* 35: 121–133.

Stack S (2009) Copycat effects of fictional suicide: a meta-analysis. In: Stack S and Lester D (eds) *Suicide and the Creative Arts*. New York: Nova, pp. 231–243.

Sudak HS and Sudak DM (2005) The media and suicide. *Academic Psychiatry* 29: 495–499.

Sundar SS and Marathe SS (2010) Personalization versus customization: the importance of agency, privacy, and power usage. *Human Communication Research* 36: 298–322.

Thaler RH and Sunstein CR (2008) *Nudge: improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thornton L, Handley T, Kay-Lambkin F, et al. (2017) Is a person thinking about suicide likely to find help on the Internet? An evaluation of Google search results. *Suicide and Life-Threatening Behavior* 47: 48–53.

Thornton P (2017) Geographies of (con)text: language and structure in a digital age. *Computational Culture: A Journal of Software Studies*. Available at: http://computationalculture.net/geographies-of-context-language-and-structure-in-a-digital-age/

Till B and Niederkrotenthaler T (2014) Surfing for suicide methods and help: content analysis of websites retrieved with search engines in Austria and the United States. *Journal of Clinical Psychiatry* 75: 886–892.

Till B, Tran US, Voracek M, et al. (2017) Beneficial and harmful effects of educative suicide prevention websites: randomised controlled trial exploring Papageno v. Werther effects. *British Journal of Psychiatry* 211: 109–115.

UNICEF (2017) *Children in a Digital World*. New York: UNICEF.

Van Deursen A, van Dijk J and Helsper E (2014) *Investigating outcomes of online engagement*. Media@LSE Working paper series no. 28. London: Department of Media and Communications, London School of Economics and Political Science.

Van Dijk J (2005) *The deepening divide: inequality in the information society*. Thousand Oaks, CA: SAGE.

Van Dijk J (2017) Digital divide: impact of access. In: Rössler P, Hoffner CA and van Zoonen L (eds) *The international encyclopedia of media effects* (pp. 1–11). New York: Wiley-Blackwell.

Van Dijk J and Hacker K (2003) The digital divide as a complex and dynamic phenomenon. *Information Society* 19: 315–326.

Võ ML-H, Conrad M, Kuchinke L, et al. (2009) The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods* 41: 534–538.

Warschauer M (2002) Reconceptualizing the digital divide. *First Monday* 7. Available at: http://firstmonday.org/article/view/967/888

Wasserman D (2016a) Suicidal people's experiences of trauma and negative life events. In: Wasserman D (ed.) *Suicide: an unnecessary death*. 2th ed. Oxford: Oxford University Press, pp. 145–228.

Wasserman D (2016b) *Suicide: an unnecessary death*. Oxford: Oxford University Press.

WHO (2014) Preventing suicide: a global imperative. Available at: http://www.who.int/entity/mental_health/suicide-prevention/exe_summary_english.pdf?ua=1

WHO (2017) Preventing suicide: a resource for media professionals: update 2017. Available at: http://apps.who.int/iris/bitstream/10665/258814/1/WHO-MSD-MER-17.5-eng.pdf?ua=1

Williams BA, Brooks CF and Shmargad Y (2018) How algorithms discriminate based on data they lack: challenges, solutions, and policy implications. *Journal of Information Policy* 8: 78–115.

Williams CL and Witte TK (2017) Media reporting on suicide: evaluating the effects of including preventative resources and psychoeducational information on suicide risk, attitudes, knowledge, and help-seeking behaviors. *Suicide and Life-Threatening Behavior* 48: 253–270.

Zillien N and Hargittai E (2009) Digital distinction: status-specific types of Internet usage. *Social Science Quarterly* 90: 274–291.

Author biographies

Sebastian Scherr (PhD, University of Munich) is an assistant professor at the School for Mass Communication Research, University of Leuven, Belgium. His research interests focus on differential media uses and effects in health communication and political communication, with a special emphasis on mental health, suicide prevention, and empirical methods.

Mario Haim is a doctoral student and research assistant at the Department of Communication Science and Media Research, University of Munich (LMU), Germany. His research interests mainly circle around effects of computation and algorithmization on journalism and media use, political communication, and health communication as well as computational methods.

Florian Arendt (PhD, University of Vienna) is a post-doctoral researcher at the Department of Communication Science and Media Research, University of Munich (LMU), Germany. His research focuses on health communication, media stereotyping, political communication, and communication research methods.

Appendix 1

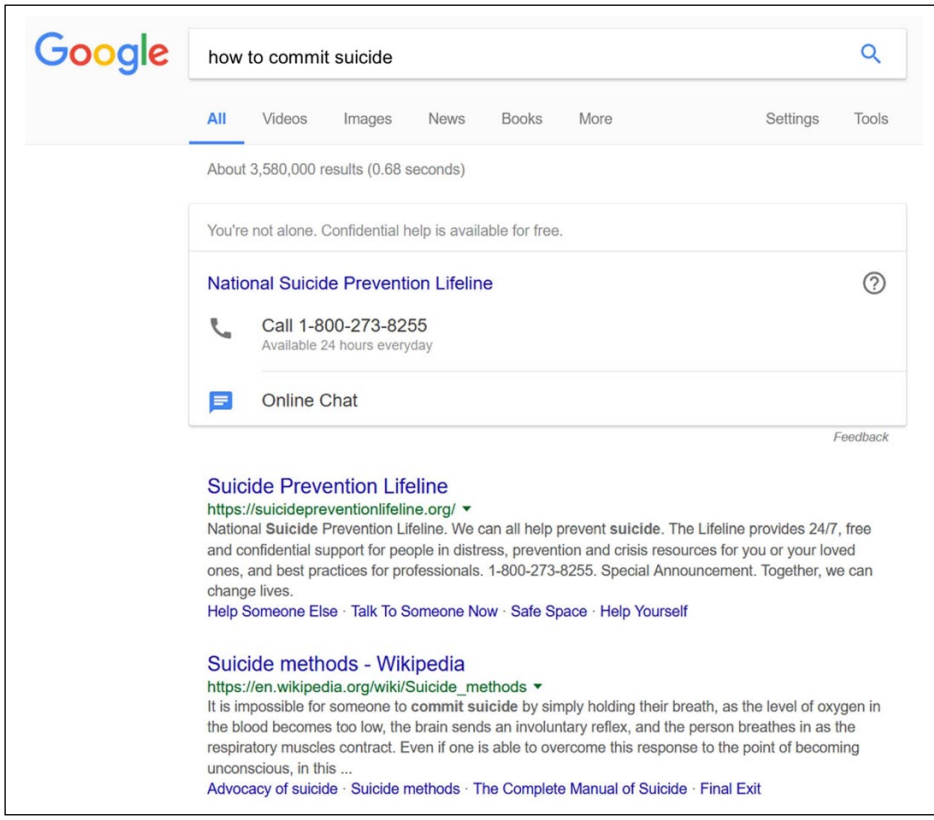

Figure 4. Example Screenshot of Google's Suicide Prevention Result (SPR).